Faculty of Science, Technology, Engineering and Mathematics
M140 Introducing statistics

M140

TMA 02 2019B

Covers Units 3, 4 and 5 **Cut-off date 9 May 2019**

Please read the Assessment Guide on the module website before beginning work on this TMA. You can submit your TMA either by post or electronically using the University's online TMA/EMA service.

This TMA is marked out of 50. Your overall score for this TMA will be the sum of your marks for each question. Note, however, that because the University requires TMAs to be marked out of 100, the mark returned to you by the University will actually be out of 100 (i.e. twice the total of marks on your TMA script).

The marks allocated to each part of each question are indicated in brackets in the margin.

Guidance about how to answer TMA questions is given in Subsection 7.2 of Unit 1.

Note that the Minitab files that you require for this assignment should be downloaded from the 'Assessment' area on the module website.

Copyright © 2018 The Open University
WEB 06683 2
12.1

You should be able to answer Questions 1 and 2 after you have studied Unit 3. You will need to use Minitab to answer Question 1.

You should be able to answer Questions 3 and 4 after you have studied Unit 4.

You should be able to answer Question 5 after you have studied Unit 5. You will need to use Minitab to answer Question 5.

Question 1 (Unit 3) – 11 marks

The Minitab file that you require for this question should be downloaded from the 'Assessment' area on the module website.

An office furniture manufacturer has trialled a new method of adhesive application. The new method is compared with the old method by performing a 'pull test' to determine the number of pounds of pressure that is required to pull apart glued furniture parts. The test results of two independent samples, each of size 22, are recorded in the Minitab worksheet **glue.mtw**. The column labelled `New method` contains the pressure (in pounds) needed to pull apart furniture parts glued by applying the new method, and the column labelled `Old method` contains the pressure (also in pounds) required to pull apart furniture parts glued by the old method. Run Minitab and open this worksheet.

(a) Produce horizontal boxplots of the data for each of the two methods of furniture gluing on the same diagram using Minitab. Make sure to include an informative label for the horizontal axis. Include a copy of your plot in your answer. [*3*]

(b) The manufacturer claims that the new adhesive method is in fact more reliable than the old method. Is this confirmed by the boxplots that you obtained in part (a)? Explain why or why not. [*2*]

(c) On the basis of the boxplots that you obtained in part (a), describe the symmetry and/or skewness of the batches of data from each of the two methods, giving reasons for your answer. [*4*]

(d) You should observe a star associated with the boxplot of the pressure required for the new method. What does this star signify? [*2*]

Question 2 (Unit 3) – 5 marks

In Unit 3, the earnings ratio of women and men is explored. Here we also look at earnings ratios of women and men in different regions of the UK. Table 1 gives the median gross weekly earnings for full-time employees by region and gender in 2017.

Table 1 Median gross weekly earnings (£) by region and gender in 2017

Region	Women	Men
North East England	452.3	545.1
North West England	464.6	550.1
Yorkshire	442.7	543.5
East Midlands	430.4	545.9
West Midlands	451.4	562.0
East England	478.4	587.5
London	622.9	766.6
South East England	504.0	621.5
South West England	459.8	563.4
Wales	457.6	525.1
Scotland	498.3	579.9
Northern Ireland	482.4	514.9

(a) Calculate the earnings ratio at the median for each region in Table 1, giving your answer as a percentage rounded to the nearest per cent. [*2*]

(b) Does the earnings ratio vary notably between the regions in Table 1? [*2*]

(c) Based on these data, comment on the gender inequalities in earnings in Northern Ireland relative to the other regions considered. Explain your answer. [*1*]

Question 3 (Unit 4) – 9 marks

This question concerns simple random sampling from the 'Sampling Department staff list' given in Table 1 of Unit 4 (Section 2). You are going to choose a simple random sample of size 14 from that list using the random number table provided in the appendix to Unit 4, starting at the beginning of row **57**.

(a) Write down the 14 distinct random numbers that you will use to identify the individuals in the sample. If you discard any of the random numbers, then explain why. [*3*]

(b) In a table, list the names of the 14 sampled individuals in the order in which they were chosen, along with their gender and occupation. [*2*]

(c) Assuming that the full staff list in Table 1 of Unit 4 is the population, comment on the representativeness of your sample with respect to:

  (i) gender [*2*]

  (ii) occupation. [*2*]

Question 4 (Unit 4) – 6 marks

A population of 2000 customers were asked about their satisfaction with a particular service. Their responses are given in Table 2.

Table 2

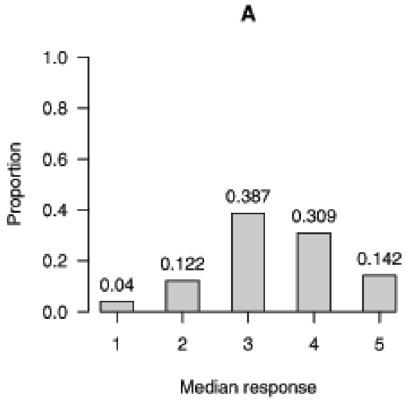
Response	Rating	Number
Very unsatisfied	1	240
Unsatisfied	2	270
Neutral	3	555
Satisfied	4	460
Very satisfied	5	475
Total		2000

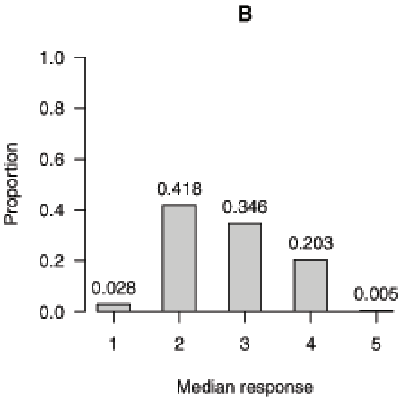
(a) What is the median response for this population? Show your working. [*2*]

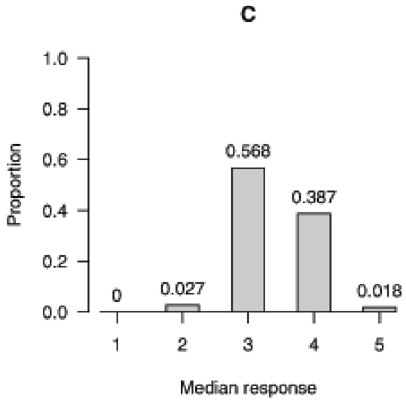
(b) Figures A, B and C below each show a distribution of the sample median, but they are based on sample medians for samples of different sizes from different populations.

- Sample 1 is the distribution of the sample median for samples of size 3 from the population described in Table 2.
- Sample 2 is the distribution of the sample median for samples of size 13 from the population described in Table 2.
- Sample 3 is the distribution of the sample median for samples of size 7 from a population different to that described in Table 2.

Giving your reasons, say which figure relates to which sample. [*4*]







Question 5 (Unit 5) – 19 marks

The Minitab file that you require for this question should be downloaded from the 'Assessment' area on the module website.

Certain environmental and medical studies are interested in measuring the average yields of tar, nicotine and carbon monoxide (CO) per cigarette, under different conditions. A subset of the data provided by one such study gives the number of milligrams of tar and CO per cigarette for 12 different brands in the USA. These data are recorded in the Minitab worksheet **smoke.mtw**. The first column, labelled `Tar`, contains the tar content per cigarette in milligrams, and the second column, labelled `CO`, contains the carbon monoxide (also in milligrams) per cigarette. Run Minitab and open this worksheet.

(a) Using Minitab, produce a scatterplot with CO on the y-axis and tar on the x-axis. Include a copy of your plot in your answer. [*1*]

(b) Now interpret the scatterplot that you produced in part (a), giving reasons for your answers and using the following checklist for interpreting scatterplots.

- (i) Is the relationship positive, negative or neither? [*2*]
- (ii) Is the relationship linear or non-linear? [*2*]
- (iii) Is the relationship strong or weak? [*2*]
- (iv) Are there any potential outliers? If so, which points are they? [*2*]

(c) Use Minitab to fit a least squares regression line to the data in **smoke.mtw**. (In doing so, you should treat CO as the response variable.) What is the equation of the least squares regression line in terms of x and y? In your answer, include a printout of the Minitab output showing your equation. [*2*]

(d) Produce a residual plot using Minitab. Include a copy of your plot in your answer. Comment on what the residual plot tells you. [*4*]

(e) You should use the equation of the least squares regression line (and hence your calculator rather than Minitab) to answer this part.

Estimate the CO content per cigarette for a brand containing 13 milligrams of tar per cigarette. [*2*]

(f) Given that, for a brand with 13 milligrams of tar per cigarette, the actual CO content was 14 milligrams per cigarette, comment on both your estimate and the general usefulness of this analysis. [*2*]